BEHAVIOR-BASED CLUSTERING FOR DISCRIMINATION BETWEEN FLASH CROWDS AND DDoS ATTACKS

Young Jun Heo, Jintae Oh and Jongsoo Jang
Information Security Department, Electronics and Telecommunications Research Institutey, Korea

Keywords: DDoS, Flash crowd, Cluster.

Abstract: We propose discrimination methods that classify cluster of traffic behaviour of flash crowds and DDoS attacks such as traffic pattern and characteristics and check cluster randomness. The behavior-based clustering consolidates packet into clusters based on similarity of observed behavior, e.g., source IPs are clustered together based on their pattern of destination port usage. The main objectives are to find way to proactively resolve problems such as DDoS attacks by detection and resolving attacks in their early development stages.

1 INTRODUCTION

The rapid development of high speed Internet has accelerated new Web applications and has in turn been driven by the popularity of those applications. But, it exhausts network and server resources such as network bandwidth, CPU, and memory. The causes of these are flash crowds and DDoS attacks.

Flash crowds are to the situation when a very large number of clients simultaneous access a popular Web site. It is a large surge in traffic to a particular Web site causing a dramatic increase in server load and putting severe strain on the network links leading to the server, which results in considerable increase in packet loss and congestion. The common examples of events are when a Web site is linked by some very popular news Web site such as Slashdot (Herman, Slashdot).

On the other hand, DDoS attacks are an explicit attempt by attackers to prevent legitimate users of a service from service usage. It overwhelms a target server with a huge amount of request packets, so as to saturate the target's connection bandwidth or deplete the server system resources to subvert the normal operation. DDoS attack was listed as the most financially expensive security accident on the 2004 CSI/FBI Computer Crime and Security Survey (Gordon). Recently, DDoS attacks have also been frequently reported that it is more sophisticated and attacker employ a large number of zombies and control them to emit requests to the target (Carl).

ISPs and network administrators make effort to protect legitimate users and their resources and cut off from their network and services. To do it, many classification approaches between flash crowds and DDoS attacks have been proposed, but it is still difficult to classify unambiguously the attack traffic from legitimate traffic. How well can the method distinguish attack conditions from normal conditions? So we use behavior-based clustering mechanism to classify two types of traffics(Kenneth).

Cluster analysis is one of the most prominent methods for identifying classes amongst a group of objects, and has been used as a tool in many fields such as biology, finance, and computer science. Recent work by McGregor et al. (McGregor) and Zander et al. (Zander) show that cluster analysis has the ability to group Internet traffic using only transport layer characteristics. One needs to develop behavioral differences between the two phenomena such as number of newer IP, request rates, packet event times, bytes per packet, bytes per burst, periodic throughput, and so on. In behavior-based clustering, traffic packets are grouped into clusters based on selected behavioral aspects, so that data with similar properties can be analyzed as a single cluster. Each resulting cluster is characterized in terms of a set of descriptive features which summarize the behavior represented in the clusters.

In this paper, we propose discrimination methods that classify cluster of traffic behaviour of flash crowds and DDoS such as traffic pattern and client characteristics and check cluster randomness.

Jun Heo Y., Oh J. and Jang J. (2009).
BEHAVIOR-BASED CLUSTERING FOR DISCRIMINATION BETWEEN FLASH CROWDS AND DDoS ATTACKS.
In *Proceedings of the International Conference on Security and Cryptography*, pages 140-143
DOI: 10.5220/0002225801400143

In the rest of the paper, we describes realted work and characteristics of flash crowds, DDoS attack in Section 2 and 3. Our approach of classification between flash crowds and DDoS attacks is presented in Section 4. Finally, we conclude our paper in Section 5.

2 RELATED WORK

There are many research to detect DDoS in the literature. He et al. proposed a mechanism to detect SYN flooding attack using Bloom filter (He). They update the client list with a Bloom filter; if a SYN request shows up on the network, they increase the corresponding counter for this client in the list; but if a SYN/ACK request comes from the same client, they decrease the number of the same counter by one.

Wang et al. use of the ratio of the numbers of SYN and FIN/RST (Wang). During a SYN flooding attack, there would be a significant amount of SYN packets, but the number of FIN/RST packets would not be as large as that of SYN packets.

Feinstein et al. develop a statistical approach to detect DDoS attacks (Feinstein). They exploit the characteristic that the distribution of source IP addresses during DDoS attacks is uniform, and detect DDoS attacks with the help of Chi-square statistics and entropy. However, the threshold of their detecting system depends on the statistical results and need to be changed in different network environments.

The main approaches focus on deal with DDoS attacks or abnormal situation of traffics without considering flash crowds. Even though some approaches take account of flash crowds, they set flash crowds as one of abnormal activities without distinguishing from DDoS attacks. Since flash crowds are caused by legitimate users, the countermeasure of server administrator during flash crowds is very different from it during DDoS attacks.

3 FLASH CROWDS AND DDOS ATTACKS

Before we classify flash crowds and DDoS attacks traffic, we understand their individual properties in this section. Although flash crowds and DDoS attacks share similar characteristics, it is of great interest to be able to distinguish them, because very different actions need to be done in rectifying these two events (Park).

Through analysis of flash crowds and other research efforts (Jung), some significant characteristics of flash crowds can be concluded, as stated below. These observations allow us to tell when a flash crowd arrives; how long (or short) a time we have to take defensive action; how different it is from a malicious attack; how we can utilize the locality of reference; and more.

- The number of clients in a flash crowd is commensurate with the request rate. This indicates that legitimate clients are responsible for the performance of a server.
- Network bandwidth is the primary constraint bottleneck. CPU may be a bottleneck if the server is serving dynamically generated contents.
- A small number of contents, less than 10%, is responsible for a large percentage, more than 90%, of requests. Moreover, the set of hot contents during a flash crowd tends to be small to fit in a cache. This property distinguishes flash crowds from attack traffic which is generated automatically by “bots”.

While studying the behavior of flash crowds, we need to identify and distinguish related but distinct phenomena to DDoS attacks. There are some ways to distinguish DDoS attacks from flash crowds.

- DDoS attackers are broad and very few previously seen clusters are involved in DDoS attacks.
- Client distribution across ISPs and networks does not follow population distribution.
- Cluster overlap which a site sees before and during the attack is very small.
- Per-client request rate is stable during the attack and deviates significantly from normal.

4 CLASSIFICATION FLASH CROWDS AND DDoS ATTACKS

What makes flash crowds and DDoS attacks different is user intention, which is hard to detect by the victim server. How can we use them to identify and separate DDoS attacks from flash events? In previous chapter, we consider what properties differentiate DDoS attacks from flash crowds? We use behavior-based clustering mechanism and randomness method. We divide incoming traffics into some clusters and compute randomness of clusters. The randomness of a cluster is more greater,

the cluster is more abnormal status. it means that this cluster occurs worm and DDoS attack.

It may monitor client that access the site and their request rates, and perform some checks on the content of packets such as number of newer IP, arrival rates, packet event times, packet inter-arrival times, inter-burst times, bytes per packet, cumulative bytes per packet, and periodic throughput samples.

4.1 Cluster

We define a cluster as a set of flows with the same values in one or several of the four keys, source IP address, destination IP address, source port, and destination port, which are typically used to define a flow. Our mechanism intends to aggregates traffic clusters. Especially, we focus on clusters with a fixed source/destination IP address because almost all abnormal traffic has either a fixed source or destination IP address. For example, packets of DDoS attacks often have the same destination IP address.

The goal of clustering is to divide flows into natural groups. The instances contained in a cluster are considered to be similar to one another according to some metric based on the underlying domain from which the instances are drawn. Clusters are flows with the same value in some combinations of these four keys and illustrate their corresponding examples are shown in Table I. These combinations show some characteristics of each cluster: flash crowds, flooding attack, worm, DDoS attack (Yan).

4.2 Classification Method Flash Crowds and DDoS attacks

Checking randomness of each cluster, we now describe the steps to distinguish flash crowds and DDoS attacks; (i) Construct the cluster of incoming traffic; (ii) Compute the randomness of each cluster; (iii) Decide normal and DDoS attacks in each cluster.

To construct cluster, we consider some combinations of IP header fields such as Table 1.

In Figure 1, we define the biggest cluster which only has the fixed source IP address/destination IP address cluster such as cluster A, define the clusters which have fixed value in two dimensions cluster such as cluster B. If we choose the higher level cluster B instead of cluster A to do aggregation, we can keep more information (source IP address and destination Port).

In four keys, the port numbers and the IP addresses have different sensitivity for the

Table 1: Combinations of four keys and some examples.

Combinations	Examples
srcIP	most worms
srcIP + dstIP	most portscans
srcIP + dstPort	Blaster worm
dstIP + dstPort	syn flooding attacks
	WWW flash crowds
srcIP +dstIP+srcPort	response from non-IP-spoofing syn flooding
srcIP +dstIP +dstPort	non-IP-spoofing syn flooding attacks
	MS-SQL server worm
srcIP+srcPort+dstPort	DNS flash crowds
dstIP+srcPort+dstPort	

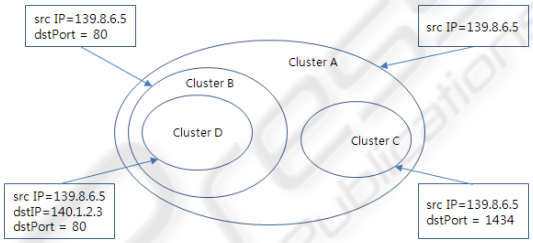

Figure 1: An example of clusters.

aggregation process. First, almost all DDoS attacks, worm spread, port scan, and flash crowds have either a common source IP address or destination IP address, but not always have a fixed port number. Second, some network applications with a well-known port number such as web traffic with port 80 are always big clusters in the network, but we have no reason to aggregate them to a single flow because they are normal traffic and we aim to maintain more detailed information about these for accounting purposes.

Besides fixed values in one or several keys, other properties of the clusters containing attack traffic include: first, the number of flows in the clusters is usually large enough to become a flooding attack; second, the size of the flows is often much smaller than normal flows; third, some keys other than the fixed value, such as source IP address in DDoS attack traffic are often randomly distributed. In addition, if there are several big flows in the identified cluster, we would pick them out from the identified cluster and do aggregation on the rest flows, because the big flows may be normal flows mixed with attack flows. In order to analyze cluster pattern, we check randomness of cluster.

We conclude several important different characteristics of flash crowds and DDoS attacks: (i) the number of requests sent to the server would increase dramatically during both flash crowds and DDoS attacks; (ii) the number of distinct clusters

during the flash crowds is much smaller than the number of distinct clients. But, DDoS attacks requests come from clients widely distributed across clusters in the Internet; (iii) a large number of clusters active during flash crowds had also visited the sites before the event. However, in the case of DDoS attacks, an overwhelming majority of the client clusters that generate requests are new clusters not seen by the site before the attack.

4.3 Experimentation

In the simulation, we use the 2000 DARPA Data Set which includes a DDoS attack run by a novice attacker (MIT Lincoln Lab, 2000). This attack scenario is carried out over five phases. In phase 1 and 2, the attacker sends ICMP packet to probe of IP's to look for the sadmind daemon running on Solaris hosts. The attacker installs Trojan mstream DDoS software on hosts in Phase 3 and 4. In Phase, the attacker launches the DDoS attack. The number of packets and randomness variation shows in figure 2 and 3.

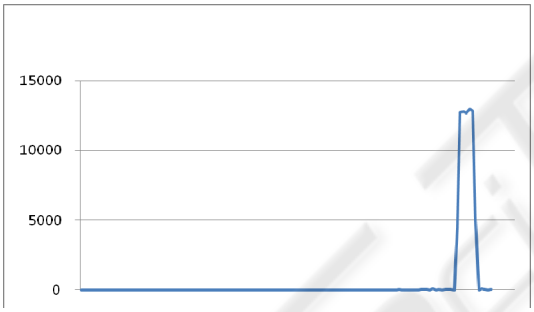

Figure 2: The Number of packets.

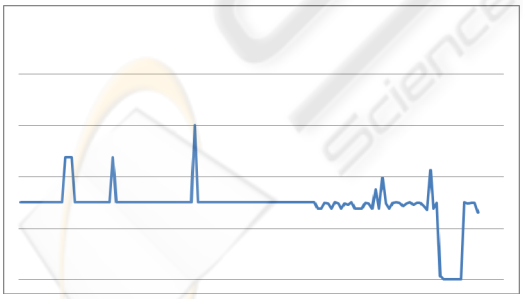

Figure3: The randomness of source IP address in Destination IP address cluster.

5 CONCLUSIONS

In this paper, we propose discrimination methods that classify cluster of traffic behaviour of flash crowds and DDoS attacks such as traffic pattern and characteristics and check cluster randomness. The main research objectives are to find way to proactively resolve problems such as DDoS attacks by detection and resolving attacks in their early development stages.

In the future work, we expect to analyze network traffic more effectively by extracting more variables and develop an advanced detection algorithm. We plan to find a way of mitigating DDoS attacks by using this early detection.

REFERENCES

U. Herman, 2006. Flash Crowd Prediction, Master's Thesis, Warsaw University.

SLASHDOT. http://slashdot.org.

Gordon, L.A., Loeb, M.P., Lucyshn, W., Richardson, R., 2004. CSI/FBI computer crime and security survey. In *Computer Security Inst..2004*

G. Carl and G. Kesidis, Denial-of-Service Attack Detection Techniques, *IEEE Internet Computing 2006*, IEEE Computer Society.

Kenneth Theriault, Daniel Vukelich, Wilson Farrell, Derrick Kong, John Lowry, Network Traffic Analysis Using Behavior-Based Clustering

Krishnamurthy, B., Wang, J., 2000. On network-aware clustering of web clients. In *ACM SIGCOMM'00*.

Jung, J., Krishnamurthy, B., Rabinovich, M., 2002. Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites. In *WWW 2002*.

A. McGregor, M. Hall, P. Lorier, and J. Brunskill., 2004. Flow Clustering Using Machine Learning Techniques. In *PAM 2004*, Antibes Juan-les-Pins, France.

S. Zander, T. Nguyen, and G. Armitage., 2005. Automated Traffic Classification and Application Identification using Machine Learning. In *LCN'05*, Sydney, Australia.

He, Y., Chen, W., Xiao, B., 2005. Detecting SYN flooding attacks near innocent side. *In MSN 2005*.

Wang, H., Zhang, D., Shin, K.G., 2002. Detecting SYN flooding attacks. In *INFOCOM2002*.

Feinstein, L., Schackenberg, D., Balupari, R., Kindred, D., 2003. Statistical approaches to DDoS attack detection and response. In *DISCEX 2003*.

Peng, T., Leckie, C., Rnmamohanarao, K, 2004., Proactively detecting Distributed Denial of Service attacks using source IP address monitoring. *Networking 2004*.

H. Park et al, Distinguishing between FE and DDoS Using Randomness Check, In *ISC 2008.*

Yan Hu, Dah-Mng Chiu, and John C.S. Lui, Entropy Based Flow Aggregation, In *Networking 2006.*